

---

ALAN MCKAY

The Queen's University of Belfast, Northern Ireland

[alanmckay33@yahoo.co.uk](mailto:alanmckay33@yahoo.co.uk)

---

# CONSTITUTION, MECHANISM, AND DOWNWARD CAUSATION

---

## *abstract*

*I develop an account of ordinary physical causation as productive, causally closed, and operating via mechanisms. This picture entails rejection of Baker's claims that intention-dependent properties are independently causally efficacious and share the lower-level physical causal nexus. However, I suggest that Baker's constitution account has the resources to overcome these difficulties, and that intention-dependent causal relations are constituted by lower-level ones.*

---

## *keywords*

*Constitution, manifest image, mechanism, downward causation*

1. **Introduction** In the final chapter of *Naturalism and the First-Person Perspective* (Baker 2013, pp. 207-234), Lynne Rudder Baker builds upon the causal arguments developed in her earlier work (e.g., Baker 2000, 2007) as part of her constitution account of reality. In that account, Baker distinguishes those objects and properties that are intention-dependent (ID) from other, lower-level, non-ID objects and properties. ID properties are either propositional attitude properties – believing, etc. – or properties whose instances presuppose that there are entities that are bearers of propositional attitudes (Baker 2007, pp. 11-13), such as the property of being an economic recession. ID objects are either such entities (i.e., persons) or objects, like houses or computers, whose existence presupposes the existence of the former. ID objects and properties are constituted, in favourable circumstances, by the lower-level, non-ID ones. However, Baker (2013, p. 217) also contends that, like all properties and property-instances, mental and other ID properties are nevertheless *physical*. It is central to Baker's anti-reductive causal arguments that ID causal property-instances are real and capable of independently causally affecting the objects and properties of the non-ID, physical world. Thus she claims that there is downward causation, whereby mental contents have physical effects, and she presents empirical data which she believes support this claim (Baker 2013, pp. 220-233). Baker's theoretical argument for downward causation is based on two claims that are, I argue, false and in any case mutually incompatible; firstly, that the causal powers of ID property-instances are independent of those of their constituting property-instances (Baker 2013, pp. 216), and secondly, that ID and lower-level causes, both being physical property-instances, belong in a single causal nexus, allowing inter-level causation (*ivi*, pp. 217; 231-233). Further, I will argue that on an account, which I will develop, of causal relations amongst the objects that make up the furniture of the everyday world, the idea that mental content, *qua* content, has effects in the physical world is incoherent. Nevertheless, I will claim, Baker's constitution account itself contains the resources to provide a robust and satisfying account of mental causation.
2. **The Constitution Account and Independent Causal Efficacy** Clarification of my proposals requires a brief review of the relevant aspects of Baker's constitution account. Constitution, according to Baker, is a relation of unity without identity, a category that lies between identity and separate existence without being either. The constitution account, which presupposes that reality contains multiple hierarchical ontological levels or layers, is developed most fully in Baker (2007) as the basis of a defense of

the reality of everyday objects and properties and their causal powers. Here I discuss only property constitution, according to which an instancing of a lower-level property in an object constitutes, in the presence of favourable circumstances, an instancing of a higher-level, for example ID, property in that object. This higher property-instance acquires, in virtue of its constitution in the favourable circumstances, novel and irreducible causal powers not possessed by its constituting property-instance alone. Favourable circumstances, in Baker's technical sense (Baker 2007, pp. 160-161), are extrinsic or relational properties that must be instantiated if the constituting property is to constitute the higher property in question. So, to introduce one of Baker's examples, an instance of hand-raising, in favourable circumstances, constitutes an instance of voting. In this case, the favourable circumstances comprise the hand-raising's being deliberately performed as a voting, by a competent person, in an environment in which there is agreement, within a suitable background cultural milieu, that a ballot is in progress in which hand-raising counts as voting. In different circumstances the same hand-raising might have constituted something else, say a call for attention, or nothing at all. Crucially, Baker insists also that the identity of the constituting thing is *subsumed* in the identity of what it constitutes. "As long as *x* constitutes *y*, *y* encompasses or subsumes *x*" (Baker 2000, p. 33), so that "*x* has no independent existence" (*ivi*, p. 46). The hand-raising *is* the voting – the "is", not of identity, composition, or predication, but of constitution. Baker's claim that constituted property-instances, such as being a voting, are endowed with novel and irreducible causal powers is encapsulated in the Principle of Independent Causal Efficacy (ICE) (e.g., Baker 2013, pp. 216):

An irreducible higher-level property-instance (*x*'s having *F* at *t*) has independent causal efficacy if and only if

- (1) *x*'s having *F* at *t* has an effect *e*, and
- (2) *x*'s having *F* at *t* would have had the effect *e* even if its constituting property-instance had been different, and
- (3) *x*'s having *F* at *t* confers causal powers that could not have been conferred by its constituting property-instance alone.

Baker (2007, pp. 115-116) offers an example in support of (ICE): Let

V be Jones's voting against Smith at *t*  
 P be Jones's hand going up at *t*

V\* be Smith's getting angry at Jones at t'

P\* be Smith's neural state at t'

C be circumstances that obtain at t in which a vote is taken by raising hands

Suppose V is constituted by P and V\* by P\*.

In the example it is assumed as a premise that Jones's voting *causes* Smith's anger. Baker's (2007, pp. 106) justification of this assumption, on the grounds of the practical indispensability of such causal claims in everyday life, is a key motivating factor in her rejection of Jaegwon Kim's arguments against non-reductive physicalism, and especially of his principle of causal/explanatory exclusion (Kim 1993, pp. 250; 1998), which states that there is no more than one complete and independent cause (or causal explanation) of any event. If Kim's arguments are accepted, Baker points out (2007, pp. 106–110), this would threaten not only the independent causal efficacy of mental content but also that of a huge range of non-mental ID properties, such as being a driver's licence or being a delegate, and for her this amounts to a *reductio ad absurdum* of Kim's position.

Baker claims that V's causing V\*, in the example, is independent of any lower-level causal relation, thus vindicating (ICE), since, first, V could have been constituted differently, for example if votes were cast electronically, and still have caused V\*, and second, although the causal powers of P alone are purely lower-level, P's constitution of V in favourable circumstances gives V the new power of causing Smith's anger.

The notion of cause that underpins Baker's claims here is a metaphysically undemanding one. Essentially, on her view, wherever a causal explanation is available and a counterfactual dependence of an explanandum on an explanans can be shown, a cause is also to be found (Baker 1993). I will now put forward an account of causal relations among the ordinary physical objects and substances that comprise our world that, I believe, calls Baker's account of ID causation into question.

### 3. Causation in the Manifest Image

Baker's rejection of the principle of causal/explanatory exclusion (Baker 2007, pp. 99–102) trades on the possibility that a fundamental microphysical causal level – the level at which true causal relations must be located, according to the exclusion principle – may not exist. I will not try to counter this argument because, I contend, this hypothetical level is not the appropriate place in which to look when we are seeking a clarification of mental and ID causation in the everyday world whose existence Baker's arguments in *The Metaphysics of Everyday Life* (2007) are aimed at establishing.

We should look, rather, at causality as it concerns the ordinary objects, with their properties and relations, that make up the perceptible, non-ID macroscopic world in which we live, together with some of its well-understood extensions into the microscopic. This is the world that corresponds to what Sellars (1991, pp. 1-40) called the manifest image of man in the world. My claim is that no matter how problematic the notion of causation may be at a fundamental level, there exist objectively real causal relations among these observable physical entities, 'objectively' being understood in Baker's (1995, pp. 232-236) sense of *recognition-independence*, in that facts about these causal relations generally do or do not obtain independently of any individual's or community's beliefs about them. Sellars himself opposes the scientific image to the manifest, and claims that the occupants of the former are the only true existents. But, as many have pointed out, this very claim, as well as all other claims, is made from the standpoint of the manifest image. Baker's argument for the reality of the world of macroscopic objects is based on practical necessity, her idea that "metaphysics should not swing free of the rest of human enquiry ... [it] ... should be responsive to reflection on successful cognitive practices, scientific and nonscientific" (Baker 2007, p. 15). Philosophers such as McDowell (2000) and Davidson (2001) further argue for a *transcendental* link between our very possession of the conceptual capacities we do and the existence of the world revealed to us through perception.

I would argue, then, that the manifest image is the natural home of our causal claims and beliefs about the world, and that it is within the manifest image that we should expect to locate the relevant distinctions among and constraints on those beliefs. We have, I suggest, a deep and intuitive understanding of what is and is not causally possible within the manifest world. We know, for example, that macroscopic objects cannot change their spatial location from *a* to *b* without passing through space between *a* and *b*. As de Muijnck puts it (2003, p. 46), if we cannot find any physical influences connecting alleged cause and effect, we would sooner suspect coincidence than "action at a distance" – that is, than some kind of magical cause-like process. I will use these notions to distinguish a basic category of causation within the manifest image that I call "manifest physical causation".

Further, I contend that our 21<sup>st</sup> Century manifest image includes objects, properties, and relations belonging to the special physical sciences, as in the biochemical example in the next section. This claim is justified, I believe, because even though such things as genes and neurotransmitters are visible only by special techniques, not only are their existence and properties so well-established empirically as to be effectively beyond doubt, but they

clearly participate in the same causal nexus as more familiar, macroscopic entities.

Ordinary causal-explanatory claims, descriptions, and explanations contain multiple instances of the use of “cause”, “because”, and their cognates which, when they cite causes and effects, move freely among mental, non-mental ID, and non-ID items. In this everyday causal discourse we usually do not distinguish either between causation and causal explanation (Beebe 2004, p. 293), or among events, states, objects, facts, or negative facts, as causal *relata*. But when we unpack this causal discourse, I will argue, we can distinguish a more basic category of causal statement. Causal claims that I categorize as manifest physical, like

a lightning strike caused the forest fire, or

local electrical depolarization of the axonal membrane causes opening of voltage-gated sodium ion channels

are distinguishable, I claim, from ID and mixed ID/physical causal claims such as

excessive sub-prime mortgage lending caused the recession,

he purposely threw the ball that smashed the window, or

human economic activity causes climate change

in a number of crucially important ways. It is important, moreover, to emphasise that our understanding of these differences is grounded in our intuitive grasp, based upon shared experience, of how things generally work in the non-intentional world around us.

Firstly, manifest physical causal statements are free of allusions to normativity or related properties that are connected with our interests, such as meaningfulness or goal-directedness. Secondly, as remarks such as de Muijnck's, above, suggest, we have every reason to think that these causal relations form a single, closed causal nexus. My inclusion of an example from neurophysiology in the category of manifest physical causation is justified, I believe, because we cannot nowadays seriously doubt the existence of such entities as neurons or axons, or that their properties are components of a single shared causal nexus, even though they are not strictly part of the world of the manifest image in its pre-scientific

form. While it is true that our understanding of special physical sciences such as neurophysiology probably does not reflect the nature of reality as postulated by fundamental physics, nevertheless within the context of the manifest image, this understanding is *homonomic*, in Davidson's (1980, p. 219) sense, with our intuitive grasp of the workings of the macroscopic world. And this understanding, applied to, say the workings of mechanical, biological, or meteorological processes, includes the tacit conviction that they proceed entirely without any influence from outside the physical causal nexus. Even when we consider human agency, whatever our view of mental causation, Tyler Burge is surely right that we do not think of mental causes "on a physical model – as providing an extra 'bump' on the effect" (Burge 1993, p. 115).

4. On my account manifest physical causation is causation in a *productive* sense. Thus when a manifest physical causal relation is instanced we understand that there must occur a transfer of energy of some kind – mechanical, electromagnetic, or chemical, say. This implies, firstly, that these causal relations are instantiated in virtue of *intrinsic* properties of the causes, and secondly, that an appropriate kind of *spatio-temporal* connection must exist between cause and effect (Hall 2004).

In contrast, the criteria by which we identify ID or mixed ID/physical causal relations are much less rigorous, being mainly based on the requirement that there be a counterfactual *dependence* of effect upon cause. Manifest physical causes, of course, also show counterfactual dependency, but the difference is that in their case the counterfactuals are grounded in properties of the manifest physical world.

Wim de Muijnck (2003) and Ned Hall (2004) acknowledge the differences between the dependence and production accounts of causation and believe that they mark an unavoidable duality in our concept of causality (de Muijnck 2003, pp. 41–50). Each of these authors independently claims that we need *both* concepts because there are some imaginable causal scenarios, such as pre-emptions, which resist analysis in terms of counterfactuals, and others, such as instances of causation by omission, that resist analysis in terms of production; thus, it is claimed, neither can provide a univocal account.

The biggest barrier to acceptance of the productive account has been the problems of causation by omission (or disconnection) and causation of omission (or prevention). For example, Schaffer argues that "causation by disconnection is causation full force" (Schaffer 2000, p. 289). The production approach cannot accommodate causation by disconnection, he claims, since

the latter “involves no persistence line between disconnecter and effect, but rather the severing of one” (ivi, p. 291). The hallmarks of productive causation, intrinsicity and *spatio*-temporal connection, are absent. Schaffer points out, for example, that when a victim is shot through the heart, the cause of death is *prevention* of oxygen from reaching the brain. I would argue, however, that this merely seems to be a case of causation by disconnection. The example is a contextual and interest-bound description of manifest physical events, framed so as to meet our explanatory needs. If we analyze the process, not as a death by shooting, but at a lower, or simpler, level of description – if, that is, we bracket our natural tendency to think of the life-death contrast as the all-important explanandum, we find we can describe the process in terms of changes in intracellular metabolism without alluding to disconnections or omissions at all. I claim that all instances of manifest physical causation are capable of description purely in productive terms.

The reason references to phenomena like omissions and preventions feature in descriptions of manifest physical causal systems is that when those systems’ physical parts are arranged in suitable ways they constitute causal *mechanisms*. Glennan defines a mechanism as “a set of systems or processes that produce phenomena in virtue of the arrangement and interaction of a number of parts” (Glennan 2009, p. 315) and goes on, “discovering a mechanism is the gold standard for establishing and explaining causal connections” (*ibidem*). There seems to be increasing recognition that the study of mechanisms, rather than discovery of laws, is an appropriate line of inquiry for the philosophy of the special physical sciences. Craver and Bechtel (2007) give an account of mechanisms in neurophysiology that emphasizes the contrast between intralevel causation and interlevel constitution. Although their notion of constitution is not Baker’s technical one, there are clear parallels; the suitable arrangement of parts might be said to be the favourable circumstances whereby an aggregate of parts constitutes a mechanism.

I claim, then, that manifest physical causation is norm-free, causally closed, productive, intrinsic, and involves the operation of mechanisms. In contrast, an ID causal relation such as Jones’s voting making Smith angry is neither norm-free, productive, intrinsic, or mechanistic in anything like the same sense, and in light of this it seems that Jones’s voting, as a higher, constituted, and *ex hypothesi* independent causal power, has no place in the manifest physical causal nexus.



5.  
Manifest  
Physical  
Causation and  
Independent  
Causal Efficacy

Baker's argument, above, for the independent causal efficacy of constituted, ID property-instances, appears valid, but depends on acceptance, on the basis of reasons that are external to the argument, of the premise that Jones's voting, V, is indeed the cause of Smith's anger, V\*. Yet I think many would agree that the validity of this premise is just what is at issue. Can the argument itself establish its validity?

Baker claims that V's causing V\* is independent in the sense that it does not depend on any lower-level causal process. But such a process undoubtedly exists; Jones's hand-raising, P, causes light rays to travel to Smith's retinae, whence neural events are initiated that lead to the instantiation of Smith's neural state, P\*. Call this causal chain or mechanism P&ae's causing P\* (ae for additional events). A causal relation between V and V\*, however, cannot be inferred from P&ae's causing P\*; nothing at the ID level corresponds to the manifest physical, mechanistic causal chain component "ae". And on the constitution account, the instantiation of P\*, caused by P&ae, guarantees, in favourable circumstances, that of V\*, so that, from the perspective of the argument, there seems to be no need for V to cause V\*.

Further, *ex hypothesi*, Jones's hand-raising, P, and his voting, V, are both physical property-instances. So V's independent, irreducible power of causing V\* must be a *physical* causal power. But in the constitution sense, V is P – it is just P in the presence of certain relational properties, which, according to the account, confer on it extra physical causal powers. If the account of manifest physical causation I have given is correct, Baker's account leaves the nature and origin of these new physical powers, and how they could be efficacious in the same causal nexus as the lower-level powers, quite mysterious.

I conclude that Baker's version of higher causal efficacy cannot work.

Her insistence that ID property-instances are physical, and hence that ID causation is of the same basic kind as lower-level causation, obscures deep differences between the two. Baker's claim that all property-instances are physical seems to be based upon the assumption that the constitution relation of unity without identity dictates that constituted entities be of the same general kind as their constituters (Baker 2007, p. 161). I think, however, that the relational qualities that ID property-instances acquire *via* the favourable circumstances of their constitution are such that to insist that these instances are physical, despite their lacking the marks of manifest physicality and causality that I have identified, is just to deprive the term "physical" of any useful discriminatory ability.

## 6. Constituted Causation

Nevertheless, I agree with the commonsense view that, say, Jones's voting does indeed cause Smith's anger. A way of protecting our ordinary intuitions about ID causation, I propose, is to claim that not just ID causes and effects, but the causal *relations* between them, are constituted by manifest physical causal relations in favourable circumstances. Thus, on this proposal, the causal relation P&ae's causing P\*, in the presence of circumstances that are essentially the same as those favouring the constitutions of V and V\*, *constitutes* the causal relation V's causing V\*. The former relation just is the latter in the constitution sense of "is", but it is transformed, in the presence of its personal and cultural relational milieu, from a mere manifest physical relation into a vastly enriched, multi-faceted ID causal relation. Further, in line with Baker's constitutional claims and our intuitions, the ID relation *subsumes* the physical one, thus vindicating our claim that it is the *real* causal relation.

ID causation, on this account, belongs in a quite different causal nexus from manifest physical causation, a nexus whose operations are constrained, not by the laws governing energy transfer or physical mechanisms, but by such factors as inference, justification, purpose, and desire. ID and manifest physical causes do not interact directly. Causation is a diachronic, purely *intra*level relation, while the physical and ID levels are connected through the synchronic relation of constitution.

This allows an alternative to Baker's (2013, pp. 220-233) interpretation of an empirical study (Anon. 2000) which found a correlation between hippocampal size and navigation experience in London taxi drivers. Baker claims that the study shows that downward causation occurs between learning, an ID property, and these physical, hippocampal changes. On my account, however, learning is constituted by other neural changes which cause the hippocampal effects, and this causal relation constitutes a purely ID causal relation between the learning and increased navigational ability.

## REFERENCES

- Anon., *Taxi Drivers' Brains "Grow" on the Job*, Internet Report, BBC News, 2000, <http://news.bbc.co.uk/2/hi/science/nature/677048.stm>;
- Baker, L.R. (1993), "Metaphysics and Mental Causation", in J. Heil & A. Mele, (eds.), *Mental Causation*, Oxford University Press, Oxford, pp 75-95;
- Baker, L.R. (1995), *Explaining Attitudes*, Cambridge University Press, Cambridge;
- Baker, L.R. (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, Cambridge;
- Baker, L.R. (2007), *The Metaphysics of Everyday Life*, Cambridge University Press, Cambridge;
- Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
- Beebe, H. (2004), "Causing and Nothingness", in J. Collins, N. Hall, L. Paul (eds.), *Causation and Counterfactuals*, MIT Press, Cambridge, Mass., pp. 291-308;
- Burge, T. (1993), "Mind-Body Causation and Explanatory Practice", in J. Heil & A. Mele, (eds.), *Mental Causation*, Oxford University Press, Oxford, pp 97-120;
- Craver, C., Bechtel, W. (2007), "Top-down causation without top-down causes", *Biology and Philosophy*, 22, pp. 547-563;
- Davidson, D. (1980), *Essays on Actions and Events*, Clarendon, Oxford;
- Davidson, D. (2001), "The Emergence of Thought", in *Subjective, Intersubjective, Objective*, Clarendon, Oxford, pp. 123-134;
- De Muijnck, W. (2003), *Dependencies, Connections, and Other Relations: A Theory of Mental Causation*, Kluwer, Dordrecht;
- Glennan, S. (2009), "Mechanisms", in H. Beebe, C. Hitchcock, P. Menzies (eds.), *The Oxford Handbook of Causation*, Oxford University Press, Oxford, pp. 315-325;
- Hall, N. (2004), "Two Concepts of Causation", in J. Collins, N. Hall, L. Paul (eds.), *Causation and Counterfactuals*, MIT Press, Cambridge, Mass., pp. 225-276;
- Kim, J. (1993), *Supervenience and Mind*, Cambridge University Press, Cambridge;
- Kim, J. (1998), *Mind in a Physical World*, MIT Press, Cambridge, Mass.;
- McDowell, J. (2000), "Experiencing the World", in M. Willaschek (ed.), *John McDowell: Reason and Nature*, Transaction Publishers, New Brunswick, pp. 3-17;
- Sellars, W. (1991), "Philosophy and the Scientific Image of Man", in *Science, Perception and Reality*, Ridgeview, Atascadero, Cal., pp. 1- 40.